Claims

We claim:

1. A method for identifying talking heads in a compressed video, comprising:

extracting motion activity descriptors from each of a plurality of shots;

combining the plurality of motion activity descriptors of each shot, into a shot motion activity descriptor;

measuring a distance between the shot motion activity descriptor and a template motion activity descriptor; and

identifying a particular shot as a talking head if the measured distance is less than a predetermined threshold.

2. The method of claim 1 further comprising:

extracting a plurality of training motion activity descriptors from a training video including a plurality of training shots, each training shot including a training talking head; and

combining the plurality of training motion activity descriptors into the template motion activity descriptor.

3. The method of claim 2 wherein the combining is a median of the plurality of training motion activity descriptors.

4. The method of claim 2 wherein the combining is a mean of the plurality of training motion activity descriptors.

5. The method of claim 1 further comprising:

normalizing the measured distance.

6. The method of claim 1 wherein the threshold is a standard deviation σ of the temple motion activity descriptor.

7. The method of claim 1 wherein each motion activity descriptor is of the form $C_{mv}^{avg}, N_{sr}, N_{mr}, N_{lr}, \sigma_{fr}$, where C_{mv}^{avg} is an average motion vector, and N_{sr}, N_{mr}, N_{lr} are short, medium and long run zero-length motion vectors, respectively.

8. The method of claim 7 wherein the distance is measured according to:

$$D(S,T) = \frac{W_{tot}}{C_{avg}(T)} | C_{avg}(T) - C_{avg}(S) | + \frac{W_{tot}}{N_{sr}(T)} | N_{sr}(T) - N_{sr}(S) |$$
$$+ \frac{W_{tot}}{N_{mr}(T)} | N_{mr}(T) - N_{mr}(S) | + \frac{W_{tot}}{N_{lr}(T)} | N_{lr}(T) - N_{lr}(S) |$$

where W_{tot} is a normalizing weight, T is the template motion activity descriptor, and S is the shot motion activity descriptor.

9. The method of claim 1 further comprising:

measuring a distance between the shot motion activity descriptor and a set of template motion activity descriptors.

10. The method of claim 1 wherein the distance is a semi-Hausdorff distance.

11. The method of claim 1 wherein the template motion activity is modeled by a discrete function.

12. The method of claim 1 wherein the template motion activity is modeled by a continuous function.

13. The method of claim 12 wherein the continuous function is a mixture of Gaussian distributions.

14. The method of claim 1 further comprising:

extracting a plurality of training motion activity descriptors from sampled frames of a training video including a plurality of training shots, each training shot including a training talking head; and

combining the plurality of training motion activity descriptors into the template motion activity descriptor.

15. The method of claim 1 further comprising:

segmenting the video into the plurality of shots using the motion activity descriptors.

16. The method of claim 1 further comprising:

retaining only talking head shots.